Biogeosciences Discuss., 9, C8347–C8362, 2013
www.biogeosciences-discuss.net/9/C8347/2013/

Biogeosciences Discussions

Interactive comment on "Soil respiration compartments on an aging managed heathland: can model selection procedures contribute to our understanding of ecosystem processes?" *by* G. R. Kopittke et al.

G. R. Kopittke et al.

vanloon@uva.nl

Received and published: 19 February 2013

The constructive and detailed comments by reviewer 3 are very much appreciated. As will be clear from our answers below, we took good advantage of the the helpful suggestions and ideas.

Some remarks or requests by different reviewers were related. So, before answering the questions and issues raised by reviewer #3, we will discuss our view on these general issues and explain our approach in making adjustments to the original manuscript.

Questions regarding to model reporting

All three reviewers made remarks about the types of model details that were reported. Reviewer 1 commented particularly with regards to details of the fittest model, such as including qq plots and information about residual variance. Reviewer 3 indicated he/she had doubts about the low RMSE values for one model (the GLMM-T model) in which the functional form was almost identical to another model (Selsted). In addition, some elaboration was requested by both reviewers on the decisions made within the modeling process or modeling details. The remarks by both reviewers were very appropriate and we have tried to deal with these points by including the requested additions in a new table (Table 4: residual variance), in a new appendix (Appendix D: diagnostic plots) and several sections within the text.

The mixed models used in this study are able to make predictions for different location conditions, depending on one's knowledge of the location for which a prediction is to be made. A paragraph has been added in the methods section which explains that mixed models can make 'location specific' predictions (treating the location factors as 'fixed effects'), or general predictions for a not previously visited location (treating the location factors as 'random effects'). The first predictive mode uses more model parameters and leads to lower prediction errors. We generated location-specific predictions with the GLMM in the original manuscript for the calibration data (where it was possible) but not for the validation data. However, due to the remark by Reviewer 3, we realize that it is better to generate non-location-specific predictions with the mixed models for both the calibration and validation data. This makes it possible to compare the RMSE across all models and calibration/validation sets and avoids confusion.

Questions related to the validation procedure and error criteria

Each of the reviewers requested an adjustment with regard to model performance criteria and calibration/validation schemes: Reviewer 1 suggested a comparison of different calibration – validation methods (e.g. cross-validation, or calibrate using all data with

BGD

9, C8347–C8362, 2013

no separate validation). Reviewer 2 suggested the calculation and application of the Akaike information cirterion for model selection. Reviewer 3 suggested that the Nash criterion be used instead of a RMSE. We think that each of these suggestions is appropriate and interesting by itself. However, including all of these additional statistics in the results and discussing them accordingly would lead to a very long manuscript with the risk of losing focus. We have therefor chosen to include an additional appendix (Appendix C) which considers a number of additional calibration / validation methods and additional error metrics. This Appendix includes a summary of the additional methods, the results from an additional model calibration (where all data is included), additional model validation (cross-validation), and two additional error criteria (the Nash criterion and a recent modification to it). The additional methods and goodness-of-fit criteria all resulted in very similar outcomes to the results already included in the original manuscript. Appendix C is referred to from the relevant places in the results and discussion sections.

As indicated by many studies (e.g. Wilmott et al., 1985; Legates and McCabe, 1999), every model performance criterion measures something different, and the choice for a certain criterion is dependent on the model aim, i.e. which aspect(s) of the system should the model especially reproduce well.

We do not have an opinion about the most suitable criterion to evaluate the prediction of yearly soil respiration (and think the best criterion is context dependent), but in our original manuscript chose to use RMSE, since this statistic has been frequently used in recent soil respiration literature (e.g. Chen et al. 2010; Migliavacca et al. 2011; Keenan et al. 2012). After calculating additional error metrics it turned out that the correlation between the various error metrics is very high and we still choose to use the RMSE as a sufficient statistic for the purposes of our study.

Chen, X., Post, W., Norby, R. and Classen, A.: Modeling soil respiration and variations in source components using a multi-factor global climate change experiment, Climatic Change, 107, 459-480, 2011.

BGD

9, C8347–C8362, 2013

Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W. and Richardson, A. D.: Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, Global Change Biol., 18, 2555-2569, 2012.

Legates, D.R., and McCabe Jr., G. J. : Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation, Water Resour. Res., 35(1), 233-241, 1999 Migliavacca, M., Reichstein, M., Richarson, A. D., Colombo, R., Sutton, M. A., Lasslop, G., Tomelleri, E., Wohlfahrt, G., Carvalhais, N., Cescatti, A., Mahecha, M. D., Montagnani, L., Papale, D., Zaehle, S., Arain, A., Arneth, A., Black, T. A., Carrara, A., Dore, S. and Gianelle, D.: Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites, Global Change Biol., 17, 390-409, 2011.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., and Rowe, C.M. : Statistics for the evaluation and comparison of models, J. Geophys. Res., 90, 8995–9005, 1985

Answers to General Comments by Reviewer #3

A) The authors present a procedure (including model selection method and model outputs integration) to estimate the age impact on annual soil respiration in heathland. Some soil CO2 efïňĆux, NEE and ecosystem respiration measurements have been performed rigorously on trenched and untrenched plots. The corrections due to the trench effect have been correctly taken into account. The experimental part of the study is well described. In the modelling part, the models tested were simple equations with variables corresponding to available data. Consequently, (i) they don't include any relationships more "processes based" and representing the processes involved in the CO2 production (e.g. temperature dependence with energy of activation,...); (ii) the variables involved could be slightly different from potential drivers of Rs, Rh or Ra (for example in this study soil water content is measured in the mineral layer when the main

BGD

9, C8347–C8362, 2013

water storage zone inïňĆuencing respiration located in organic layer; pg 16272 L 9-11). When models are chosen before the design and set up of the experimental device, the variables represented in the selected models can be measured or determined with a higher accuracy. All this makes the simplest model to be the more representative of the data and it is unable to reproduce some situations depending of variable nontaken into account in the selection procedure (high soil CO2 efïňĆux on 21 March 2012 with, apparently, dependence to active microbial biomass). We fully agree with this this description of the study, and also agree with the reviewer that applying a model-based approach for experimental design leads to more representative, more accurate and more precise observations to enhance or test that model. In this context we would also like to bring up, that one needs to choose or develop an adequate model to start with and only when having a model which represents the system-parts of interest well, it would be effective to plan measurements or experiments based on this. So at the start of a field research one would typically observe the 'standard variables' or ones that have been successful in comparable studies according to a design based layout (we were in our field research in this initial phase). By analysing the data (and model results) from this initial phase, more targetted observations can be collected in subsequent field trials. As a result of the model analysis in this study, we will in future field trials focus on soil moisture observations in the organic layer.

B) In addition some Rs parameterisations chosen give apparently Rs dependence in opposition with some well known behaviour (see speciïňĄc comments Pg 16526, L 2-Table 3). We think the reviewer has misread the equation, see our answer to specific comment 9.

C) The way to select the "best" model is questionable, some criterion being more adapted that the RMSE, like the Nash criterion (Nash and Sutcliffe, 1970), to estimate the goodness of a model outputs comparing to a data set. This comment relates to remarks also made by Reviewer 1 and 2. We have discussed them jointly within the general section of this review-answer. We agree with the reviewer that the Nash-

BGD

9, C8347–C8362, 2013

Sutcliffe coefficient (E) is (along with RMSE) an often used goodness-of-fit criterion. And we also agree with the reviewer that it is, in general, good practice to include more than one criterion to measure model performance, as different criteria measure different aspects of model performance (see e.g. Legates and McCabe, 1999). So we have now emphasized this point in the discussion and as Appendix C. In addition we have added E as well as a modified version by Wilmott et al. 2012 in Appendix 3. However, we would like to make two side-notes to this point of adopting additional goodness of fit criteria. First of all, while we agree that the use of a different criterion may in theory lead to different results, it turns out that in our study, RMSE and E are highly correlated (Pearson correlation of -0.97, calculated on the basis of Table C3 – Appendix 3 in the revised manuscript). So the use of either RMSE or E (or the modified version by Wilmott et al, 2012) does not really make a difference. Secondly, we think it is important to use goodness of fit criteria which are commonly used and (through that) are meaningful for the 'soil-respiration community'. We think that the RMSE is by far the most used (and best understood) metric in this domain, so that is why we have chosen to use it in our original manuscript. Since we have included Nash-Sutcliffe coefficient (E) in Appendix C of our revised manuscript, we take the opportunity to briefly provide some background information about this coefficient. E belongs to the class of dimensionless goodness-of-fit criteria of the form F = 1 - d/μ, where d is a dimensioned measure of average error and μ is a basis of comparison (Wilmott et al., 2012). The selection of d determines which average error-magnitude will be represented, while the choice of μ determines the lower limit of F, as well as the sensitivity of F to changes in d. As d = 0 and μ > 0, the upper limit for F is 1.0 and indicates perfect model performance. In most cases, μ is defined such that the lower limit of F is 0, -1, or -ïĆě . E can range from -ïĆě to 1. There exist various modifications to E, such as the modified and relative E (Wilmott et al. 1985) and the parameter-corrected E (Clarke, 2008). We are unsure what the reviewer means by the statement that E is more adapted than the RMSE. Probably the reviewer refers to the fact that E is scaled, and that a general qualification is sometimes used to express model fit on the basis of E (0.8<E<1.0 =

BGD

9, C8347–C8362, 2013

Excellent; 0.6<E<0.8 = Good; 0.3<E<0.6 = Reasonable; 0<E<0.3 = Poor; E<0 = Bad). However the weaknesses of E have been shown and discussed in the literature. It has been argued that E (by itself) does not provide a reliable basis for comparing the results of different case studies (e.g. Schaefli & Gupta, 2007). So while we do not want to argue that E is less useful than any other goodness of fit criterion (like RMSE), we want to emphasize that it also has its limitations.

Clarke, R.T.: Issues of experimental design for comparing the performance of hydrologic models. Water Resour. Res.44(1), 1944-7973. 2008

Legates, D.R., and G. J. McCabe Jr. G.J. : Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation, Water Resour. Res., 35(1), 233-241. 1999

Schaefli, B., Gupta H.V.: Do Nash values have value? Hydrol. Process. 21, 2075–2080. 2007

Willmott, C. J. : On the validation of models, Phys. Geogr., 2, 184–194. 1981

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell J., and Rowe, C.M.: Statistics for the evaluation and comparison of models, J. Geophys. Res., 90, 8995–9005. 1985

Willmott, C.J., Robeson, S.M., Matsuura, K.: A refined index of model performance. Int. J. Climatol. 32: 2088-2094. 2012

D) In this selection procedure, it seems that two models are identical (Selsted-T and GLMM-T, pg 16289) but give very different RMSE results for the calibration (Fig. 8) This decreases the credibility of the model selection procedure. We should have explained this point better. The mixed-model used in our study contains an 'intercept-term' which varies per measurement location. The average over these 'location-specific-intercepts' is zero (or very close to zero). If you apply the mixed-model to locations that you know (as in the calibration-case), you can in principle use the intercept-term for that location

BGD

9, C8347–C8362, 2013

(you treat the location as a 'fixed factor'). But if you apply the model to generalize (predict for unknown locations, as in the validation case), you cannot use the intercept terms (you treat the location as a 'random factor'). So, because we included the intercept terms when predicting for the calibration locations, the RMSE was lower in that case. In the revised version we explained this issue better in the method-section and decided to treat the random part of the mixed-model as a random term in both calibration and validation. E) The main important conclusion of the study is the age effect on the annual carbon exchanges in heathland. For the total and autotrophic respiration ïňĆuxes (Rs and Ra), this effect is established from data and ïňĄgure (Fig. 11) without any representation of the uncertainties. Without proof of a signiïňĄcant difference between the Rs or Ra values presented, it is very imprudent to conclude to age impact. The main uncertainty due to the one on the model parameters determination (R0 and k in R=R0 exp[k*Tsoil]) should be at least proposed. The same remarks could be made for the percentage given for the ratio Ra/Rs. We agree with the reviewer that an estimate of prediction uncertainties is required to support any statements on Age-effects on Rs. In the revised manuscript, we calculated the 95% confidence intervals (using a bootstrap procedure with 1000 replications) to infer the observed significance level of the treatments. The outcomes of this bootstrap process supported the original statements that the Young vegetation Rs was greater than Rs on either the Middle or the Old communities. These confidence intervals have been included in the annual C loss figure (previously Figure 11) and the p-values have been provided in the result section to support the statements on age-effects on Rs.

Answers to SpeciïňĄc Comments by Reviewer #3

1) P 16242, L 6-9: Carbon in live roots is a pool of soil C (could be large in forest). The fact is that Rh and Ra can have very different dependence to biotic and/or abiotic factors and each of them has to be estimated separately for long term The authors agree that carbon in live roots is a pool of soil C which could be large in forests, and agree that the Ra and Rh can be differently affected. Therefore, this sentence has

BGD

9, C8347–C8362, 2013

been amended to clarify the meaning.

2) P 16242, L 17: “Once ïňĄeld data has been collected, the interpretation of the RS, RA and RH data has generally been undertaken through a comparative analysis and discussion of the original observations”. Could you clarify? In many cases, investigations have used only their original field observation data to compare differences between systems and have not used the data for modeling (such as estimation of annual C loss or forward/backward predictions of C losses). Where modeling is being used, it was historically most often used to generate and compare Q10 values. This paragraph and the following paragraph have been amended to improve readability (third paragraph in the introduction section).

3) P 16243, L 10-11: “measures of ïňĄt for the calibration data” Do you mean parameter representing goodness of the regression? Yes that is what we mean. In this context we would call it rather an ‘index’ or ‘statistic’ to represent the goodness of the regression (we like to reserve the term ‘parameter’ for constants in a model, associated with variables and which may be known apriori or derived through calibration).

4) P 16246, L 24-25: Why Untrenched Validation plots are located only in Old vegetation and not spread over the three age communities? Can this impact the validation? The Untrenched Validation plots were only located in the Old vegetation, as plots associated with a concurrent trial were already established and could easily be measured and incorporated into the three-community-age study. In an ideal situation, where resources weren’t limited, it would have been nice to collect respiration measurements from more locations within each treatment on more occasions within the study period. However, unlimited data collection was not possible and data measurement locations had to be maximized for the resources available. Collection of more data may have allowed for finer model resolution and more accurate predictions however, the outcomes of this study provide a solid base for further study and assessment of additional variables.

BGD

9, C8347–C8362, 2013

5) P 16249, L 3: “A loess smoother curve”. Could you clarify? A loess (smoother) curve is a smooth non-parametric curve, indicating the overall trend through a data set. It uses locally weighted polynomial regression were at each point in the data set a low-degree polynomial is fitted to a subset of the data, with explanatory variable values near the point whose response is being estimated. The method was originally introduced by Cleveland (1979) and has become very popular. It has been implemented in the R statistical program as well as in many other software environments. We think it is such a general and widely applied technique that it does not need extra clarification in the manuscript. Cleveland, W.S. : Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 74 (368): 829–836. 1979

6) P 16251, L 23- P 16252, L 3: Indicate here that this CO2 ïňĆuxes event is linked to a special meteorological episode (freeze followed by thaw). Could this kind of extreme events be more frequent in the future in your heathland location? If yes, it becomes important to be able to reproduce it in the model (see general comments). There is agreement on the fact that impact of extreme events will become as important as long term drift (see Carbo-Extrem program founded by the FP5 of the European Community). Your choice is clearly to study only the inïňĆuence of "uniform" climatic change. This should be taken into account with better emphasis in your discussion and conclusion (and especially they have to be more cautious). This paragraph has been updated to highlight that these extreme values were most likely associated with an extreme meteorological episode (freeze followed by thaw) that occurred in late winter, where an extreme freeze period (-20ËŽC) was followed by daytime air temperatures which rapidly reached >15 ËŽC. This paper did not aim to assess the impact of climate change on annual C loss but rather aimed to assess the difference between three heathland communities of different ages in a one year period. However, the authors agree that the frequency and intensity of extreme events such as drought and heat waves have increased in Europe and are likely to continue to do so in the future (IPCC, 2007; Dai 2011). Within northern Europe, warming extremes may be more of a concern than cold extremes as evidenced from recent trends (1961-1990) where the annual num-

BGD

9, C8347–C8362, 2013

ber of warm extremes of the daily minimum and maximum temperature distributions increased twice as fast during the last 25 years than expected from the corresponding decrease in the number of cold extremes (Klein Tank and Können (2003) reported in IPCC, 2007). However, if extreme freeze periods were to increase in frequency, soil C loss is likely to be influenced both through suppression of C loss during the frost period and subsequent elevation of C loss during the following warm periods. Additional research that focused specifically on extreme events would be required to investigate the potential effects of these extreme events on the Oldebroek heathland. IPCC: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. 2007 Dai, A.: Drought under global warming: a review. Wiley Interdisciplinary Reviews: Climate Change 2:45-65, 2011.

7) P 16252, L10: How do you "identify" an effect before t tests? An ANOVA was first used to indicate if there was any effect of the age treatment on soil respiration. Once a treatment effect was identified, then the t-test with bonferroni correction (also known as a 'post-hoc test') was used to further investigate the differences between treatments.

8) P 16252, L 26: Is not clear for which plots and communities measured or modeled values of soil moisture are used in the model calibration and validation. If measured data exist is preferable to use them in the calibration-validation process otherwise your introduce additional uncertainty due to your hydrological model (determination of the parameters,...). Model values should only be used when measurements are missing. Soil moisture and soil temperature were recorded at 5cm below ground surface on an hourly basis in two uncut plots per vegetation age and two cut plots per vegetation age, as described in Section 2.3. Therefore, continuous data was not available for all plots. A moisture model was chosen for two reasons (as described in Section 2.7.2). Firstly, a dynamic model is an appropriate method to integrate the soil moisture values per sensor to an average soil moisture value per treatment. This integration is necessary because not all plots were equipped with a soil moisture sensor (as mentioned

BGD

9, C8347–C8362, 2013

above). Secondly, it overcomes problems of missing data, such as when a respiration model is used at other sites for predictive purposes, the soil moisture data is usually not available, whereas daily rainfall and temperature are commonly present. The soil respiration model was applied to datasets in which both the original measured soil moisture data and the modeled soil moisture data was included. The model outcome was very similar for both soil moisture datasets. Therefore, based on these advantages described above, the soil moisture model was chosen for use within the soil respiration modeling.

9) Pg 16526, L 2-Table 3 : There is a problem with some functions like Rs =R0+ kT + a (M − 1) with Rs decreasing when M (=SWC/SWC at ïňĄeld capacity) rises toward one, so when SWC increased. It's in opposition with what is usually observed during drought (Rs decrease with SWC) The reviewer is right that a relation RS = R0 +kT +a(M −1) would be strange. However, the functional form (which is used in the LMM2 and GLMM2 model) is a(M −1)ˆ2 (with (M-1) squared). Since the maximum value that M can get is 1, this part of the function describes in fact a parabolic increase when M increases from 0 to 1.

10) Pg 16256,L 6: PPFD can be used as substitute for P (Photosynthesis) but not for Tsoil which is not often correlated to PPFD The authors recognize that the reviewer is referring to photosynthetic photon flux (area) density (PPFD), of which PAR is a measure. The substitution of PAR for P and Tsoil was explorative in nature and reference to Tsoil has now been removed from the sentence.

11) Pg 16258, L 9-10: Are total respiration in Middle and Old communities signiïňĄcantly different from zero in spring? (and in autumn or winter)? Yes. Soil respiration was significantly greater than zero in all seasons. A sentence has been added to the results to specifically mention winter as this may be the season which might be of most interest.

12) Pg 16260, L 17: What means signiïňĄcant for a model parameter? Could you

BGD

9, C8347–C8362, 2013

clarify? With the term parameter significance we follow the conventional definitions ('the probability to committ Type I error'). It means that the model parameter is different from zero, assuming with a 0.05 significance level.

13) Section 3.4-3.5-3.6: Make a table giving the RMSE (for calibration and validation) and including only the models and variables set for which the parameters are signiïňĄcant. Then, reduce drastically the sections 3.4, 3.5 and 3.6. Don't need to speak about the no signiïňĄcant parameter cases. A table giving RMSE for calibration and validation values for selected models has been included in the revised manuscript. The length of sections 3.4, 3.5 and 3.6 is now reduced.

14) Pg 16264, L 20-25: An estimation of the uncertainty on the percentage and annual values presented should be given (see General Comments). This has been undertaken as requested. The 95% confidence intervals are now included in the annual C loss figure (previously Figure 11) and the p-values, which were generated from a model comparison, are now provided in the text.

15) Pg 16264, L 20-25: What is the annual carbon loss estimates for the Middle age community? The annual C loss for each community was Young: 649 g C m-2 year-1; Middle: 462 g C m-2 year-1; Old: 435 g C m-2 year-1 as represented in the plot of annual C loss (previously labelled Figure 11). However, to be consistent, the Middle community values have now also been included in the written text of the results.

16) Pg 16265, L 11-13: Which the parameter set (Young, Middle or Old community) has been used to obtain 350 gC m-2 yr-1? What's represent the 322 gC m-2 yr-1? The Rh model was calibrated on all cut plots, regardless of community age. Therefore, only one set of model parameters was generated for later use with the predict function. When the respiration predictions were generated, the soil temperature input data was from an uncut plot on the Middle community. This soil temperature data was considered representative of the conditions underneath a vegetated plot and was therefore applied in the prediction of Rh on a vegetated plot. For soil respiration on a bare plot, the same

BGD

9, C8347–C8362, 2013

Rh model parameters were used but the soil temperature input data was from a cut plot on the Middle community to represent the more variable temperatures associated with bare ground. Therefore, the difference in Rh C loss on bare ground (350gC m-2 yr-1) and vegetated ground (322 gC m-2 yr-1) is associated with the temperature difference between cut plot and uncut soil temperature, as defined in the last paragraph before the discussion.

17) Pg 16266, L 20-29: Could be suppressed This paragraph has been removed.

18) Pg 16267, L 9-10: Have you some LAI data to support this argument? This section has been amended to clarify the statement about plant activity as no LAI data is available.

19) Pg 16267, L 20: "the ratio of moss biomasses and the ratio of photosynthetic rates". This data should be presented in the Results section This sentence has been updated with more recent information on Calluna PG (based on preliminary trials). This new sentence reads "This study did not quantify the separate PG contributions of moss and Calluna to the overall photosynthetic rates. However, based on the preliminary data from in a trial in May 2012, the Young Calluna plants were approximately 2.5 times more photosynthetically active than the Middle and Old Calluna and therefore, PG would still provide a measure of the plant activity for each community." It is not considered appropriate for this information to be presented in the results section.

20) Pg 16267, L 20 - Pg 16267, L7: The authors pretend that this CO2 emission peak as no impact on annual value (reason why this data was deleted from the calibration data set), but which peak duration have been chosen to draw this conclusion? How long are the periods between the 21 March 2012 (peak date) and the date of the preceding and following measurements? If peak is as long as these periods added (worst case, unrealistic but what is the real period?), is it impacting the annual C loss? The dates in which soil respiration were measured at Oldebroek around the extreme freeze and thaw period were 9 March 2012, 21 March 2012 and 30 March 2012 (for calibration

BGD

9, C8347–C8362, 2013

and validation-trenched data) and 8 March 2012 and 28 March 2012 (for validation-untrenched data). Of these, the extreme CO2 flush values were only observed on 21 March 2012. Therefore, there were 18 days between non-flush measurement dates (9 to 28 March 2012). To estimate a very broad level of error for these 18 days, we calculated the additional C loss that could have been attributed to the recorded high-flush values. These calculations assumed that the elevated flush continued during the warmer 'daytime' hours (8am to 7pm) of all 18 dates, using the mean flux rate measured on 21 March 2012 (Young: 3.05, Middle: 1.44, Old: 1.32 μmol CO2 m-2 s-1). Therefore, it was calculated that the total annual C loss would increase by 3% for Young (from 650gC to 670 gC m-2 year-1), by 1.5% for Middle (from 462 gC to 469 gC m-2 year-1) and 1.4% for Old (435 to 441 gC m-2 year-1). It is likely that these broad calculations have overestimated the impact of the flush period as the mean values measured on 21 March 2012 would not have continued at this rate for the entire period between 8am to 7pm for the 18 days. However, this estimate indicates that there is only a relatively small impact on the overall annual C loss (<3%) and the authors do not believe that these C increases would change the finding of the study that Young vegetation has greater soil C loss than the Middle or Old vegetation.

21) Pg 16271, L20-25: Repetition could be suppressed. This section has been shortened, and repetition removed.

22) Pg 16272, L 9-11: Could you use your water bucket model to estimate the soil moisture content in the organic layer? We think it could be done but we currently lack the observations (soil moisture in the organic layer, coupled to the underlying mineral layer) to do so. We do plan to collect soil moisture observations in both the organic and mineral layers in order to obtain more appropriate soil moisture observations as well as a better soil moisture model as input to soil respiration calculations.

23) Pg 16276, L19-20: How do you pass from soil respiration annual estimates to total C exchange for the ecosystem? This section has been clarified to indicate that these total C flux statements are hypotheses based on the RS and PG data used in the

BGD

9, C8347–C8362, 2013

current study. These hypotheses could be tested with the NEE and ER data using a similar modeling approach as described in this study.

Please also note the supplement to this comment:
http://www.biogeosciences-discuss.net/9/C8347/2013/bgd-9-C8347-2013-supplement.zip

---

Interactive comment on Biogeosciences Discuss., 9, 16239, 2012.